



COMP 6721

Applied Artificial Intelligence Project Report

Implementation of Naïve Bayes Algorithm used for Email Classification

GROUP: HA-G10	
Full Names	ID
Che Shao Chen	40055392
Ufuoma Ubor	40072909
Yunan Zhang	40105680

Submitted to: Dr. Rene Witte

- **Analysis**

An important validation factor in Machine Learning classification is the performance of the model, a good model must have a certain acceptable level of accuracy. Using only accuracy for performance measure is not sufficient enough as it can sometimes be misleading, e.g. a model might be very bad in predicting a certain category but its accuracy can still be very high. For this reason, it is advisable to not fully depend on accuracy but rather other metrics like precision, recall and f-measure. Table 1 below display the performance for classifying spam and ham respectively using the Naïve Bayes algorithm we implemented from scratch.

	Spam	Ham
Accuracy	0.89	0.99
Precision	1.0	1.0
Recall	0.89	0.99
F1-measure	0.94	0.99

Table 1: Performance results for classification

After prediction, the predicted labels were separated into the two categories with respect to the expected output labels. After the separation we compute the accuracy, precision, recall and f1-measure for each category. Table 2 and 3 shows the confusion matrix for both spam and ham respectively. The classifier was able to correctly predict 356 test samples as spam (true positive) and misclassified 44 as ham (false negative).

	Expected: Spam P	Expected: Ham N
Predicted: Spam P	356	0
Predicted: Ham N	44	0

Table 2: Confusion matrix for spam category

	Expected: Ham	Expected: Spam
Predicted: Ham	394	0
Predicted: Spam	6	0

Table 3: Confusion matrix for ham category

On the other hand, the classifier correctly predicted 394 test samples as ham and wrongly predicted 6 as spam.

References

- [1] Aisha, J. (2018, August 26). Unfolding Naïve Bayes from Scratch.
- [2] Praveen, D. (2018, December 18). An Introduction to Bag of Words And How To Implement it in Python for NLP
- [3] <https://stackoverflow.com/questions/12995937/count-all-values-in-a-matrix-greater-than-a-value>